Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 141 (2014) 252 – 259

Procedia
Social and Behavioral Sciences

WCLTA 2013

# Comprehensive Assessment of Teaching Performance in Medical Education

Urrutia-Aguilar ME, Sánchez-Mendiola M, Guevara-Guzmán R y Martínez-González A.

*Faculty of Medicine, Universidad Nacional Autónoma de México.*
*Faculty of Medicine, Universidad Nacional Autónoma de México.*
*Faculty of Medicine, Universidad Nacional Autónoma de México.*
*Faculty of Medicine, Universidad Nacional Autónoma de México.*

**Abstract**

Assessing teaching performance is important for students' learning. Objective: Assess teaching performance in a medical school, using three strategies. Method: 288 Teachers were studied. The three assessment strategies were: students' academic achievement measured with validated exams, teachers' evaluation by students' opinion (2,656 students) with a validated instrument, and a self-assessment tool. Results: There were significant differences ($p \leq 0.001$, ANOVA) among the three assessment strategies. Four different teacher categories were obtained using students' opinion. Conclusion: We identified high-performance teachers, those requiring faculty development and those who do not fulfill appropriately their teaching responsibilities.


Selection and peer-review under responsibility of the Organizing Committee of WCLTA 2013.

*Keywords:* Teaching evaluation; self-assessment; academic achievement; undergraduate medical education;

## Introduction

Teaching in the field of medicine faces new challenges in recent times. Just mastering his own field of knowledge is not enough for a teacher. He/She must have basic knowledge of other disciplines, knowledge of individual and group psychology, as well as Medical Humanities, and be able to use the scientific methodology and establish effective interpersonal communication (Martínez et al., 2008).

Changing the way one teaches is difficult; some teachers reproduce the model they lived as students (Sanmartí, 2004). If the aim is for students to achieve meaningful learning, teachers must be trained in effective strategies based on constructivist models. Ideally, training should be based on each teacher's needs, derived from reflection and actions that lead to a continuous improvement of the teaching-learning process. Teaching assessment should be carried out through various strategies, to provide teachers with feedback about their strengths and opportunity areas (Marsh, 1984; Irby, 1983; Fernandez, 1997).

There are three main reasons for assessing teaching performance. First, it provides important information on which

Corresponding Author: Urrutia-Aguilar ME
E-mail: urrutia-aguilar@gmail.com


Selection and peer-review under responsibility of the Organizing Committee of WCLTA 2013.
doi:10.1016/j.sbspro.2014.05.044

decisions can be made for teacher training. Secondly, it allows the teacher to develop a plan to correct deficiencies, and consolidate those activities carried out correctly. And they can support the comprehensive evaluation made by the institution and external auditing agencies. Different models are used according to the purpose and context for assessing teaching (Martínez, Fortoul & Urrutia, 2011; Martínez et al., 2011): a) the foreboding model assesses skills and attitudes that teachers have in their teaching practice (García, 1984; Gimeno & Pérez, 1989), b) the process model evaluates a set of basic competencies of the teacher's behavior in his educational space (Garcia, 1984; Ibanez, 1990), c) the product model which estimates learning results (Vega, 1986); and, d) the mediational model which articulates the product-process models (Gimeno & Pérez, 1989).

Student assessments have proved to be a consistent, valid and useful source of information for feedback to improve teaching (Marsh, 1984; Marsh, Fleiner & Thomas, 1975; Wilson, 1986). The evaluation through students' opinion is the most used and studied in North America, Europe, Asia and Mexico (Lancaster, Ross & Smith, 1988; Luna & Torquemada, 2008; Stolnick, Rusell & Durkovich, 1975; Mullan, Wolf & Ertel, 1989). We have also carried out teaching self-assessment, consisting of judging own's teaching, from various instances and based on certain assumptions looking for specific objectives (Solabarrieta, 1996; Barbier, 1993; Rippey, 1981; Urrutia et al., 2008).

Another approach to teaching performance evaluation is the students' academic achievement, considering that one of the main functions of the teacher is the promotion and improvement of students' learning. Through his/her roles as an informer and a trainer, the teacher seeks students' knowledge learning, development of skills, attitudes and values necessary to achieve professional competence (Martínez, Fortoul & Urrutia, 2011; Martínez, et al., 2011; Rodriguez, et al., 2000).

This becomes more important in universities, especially in the field of medical education, considering that the aim of these institutions is the training of future physicians who will solve Mexicans' health problems. Curricula in most Mexican universities, particularly at UNAM, are divided into yearly courses, in which during the first two years, students attend classrooms, laboratories, community health centers; and in the following three years, hospitals and ambulatory care centers. A solid academic background in basic, clinical, and sociomedical sciences in the first two years is essential to develop clinical competence, and research on teaching performance and its impact on students' academic achievement becomes necessary. Research in this area is limited in our country (Cruz, 1985; Micu, Castillo & Díaz, 1986; Talayero, 1988; Rodríguez et al., 2000). The goal of this paper is to evaluate teaching performance during the first two years of medical school in the morphological, physiological and socio-medical areas, using three strategies.

## 2. Method

Longitudinal, comparative and paired study. The study sample consisted of 288 teachers representing 81% of the total that teaches six courses in the first year (145 teachers) and five in the second year (143 teachers), at UNAM Faculty of Medicine MD program in Mexico City, during the period 2007-2008. Teachers accepted voluntarily to participate. Only teachers with an assessment by student's opinion, self-assessment and academic performance of students were included to integrate a mediational model. The courses of Anatomy, Cell Biology, Developmental Biology and Microbiology and Parasitology were grouped in the Morphological area; Physiology, Pharmacology, Biochemistry and Immunology in the Physiological area; Public Health I and II, and Medical Psychology into the Socio-medical area.

To assess teaching performance according to the students' opinion, a validated instrument was used (Valle et al., 2004) developed by the university Educational Evaluation Department (Dirección General de Evaluación Educativa), consisting of 30 items. The questionnaire had a Cronbach's $\alpha$ of 0.93; each item had five possible Likert-type answers (1=never, 2=seldom, 3=sometimes, 4=often, and 5=always). This instrument considered three dimensions: Teaching Strategies, Respect for students, and Learning Assessment. It was answered by students before they knew their course grade and participation was voluntary. 9,152 (73%) questionnaires were included.

Self-assessment was conducted by teachers voluntarily, answering an instrument based on the same categories and items used to evaluate students' opinion. Instrument's content validity and prediction of good and poor teaching performance was assessed by experts before its application (Urrutia *et al*., 2008).

The Entrance Exam Scores (EES) and the Average Graduation Exam Scores (AGES) were considered to determine the level of academic performance obtained by students. For first year medical students, PAI was the average obtained in the Diagnostic Test (designed and validated by the Department of Educational Evaluation and implemented by the Faculty of Medicine to all students at admission), used in studies of school lag or dropout (Tellez et al., 1990; Martínez et al., 1999); for second year medical students, the final average score obtained in the

partial departmental summative examinations of the previous school year courses was used to calculate PAE. It is important to highlight that three to five departmental exams are applied for each course (each composed of 50 to 70 items selected from an item bank developed and validated by content experts). The instruments had the following parameters: Cronbach $\alpha \geq 0.80$, difficulty level range = 20-80 and positive discrimination $\geq 20$.

In order to control the advantage-disadvantage factor that is based on value-added models that evaluate the impact of some variables on the teaching-learning process using statistical techniques such as ordinary least squares regression, adding the value of some variables in a process where multiple factors that interfere in students' academic achievement. This study considered the regression beta coefficient as statistical predictor (Ballou, Sanders & Wright, 2004; McCaffrey et al., 2003). The advantage-disadvantage factor given by the differences in students' EES was used to adjust their AGES by means of the regression coefficient. It was considered that the beta coefficient of simple linear regression should be used as a statistical significant predictor of the dependent variable (at graduation) from the independent variable (at entrance). Under this premise, the equations corresponding to each group were obtained by subtracting from the initial advantage or adding to the initial disadvantage in the average at graduation.

We carried out a subsequent analysis of distributions and differences of averages obtained by the students constituting the study population on entrance and at graduation (EES vs. AGES); only the cases that met the requirement of having a score in all partial and diagnostic examinations were considered. The results were analyzed with Statistical Package for the Social Sciences (SPSS) (version 19), the instrument scale from 1 to 5 was converted into percentage averages to homogenize them with the average at entrance and at graduation; descriptive statistics, data normality analysis, Student's *t* test for related samples, $X^2$, ANOVA and Pearson correlation were applied.

## 3. Results

In the study population (288 teachers), 57% were male and 43% female. The mean age was 54 years, ranging from 31-84. 75% of them were physicians and 25% belonged to other disciplines (mainly biologists and chemists). In terms of their academic level, 79% had a postgraduate degree: specialty, master or doctorate.

In general, students were satisfied with their teachers' academic performance. Teaching strategies were the factor with the least satisfactory evaluation ($84.68 \pm 11.34$). In relation to the factor called "respect for students" there were significant differences of $p \leq 0.05$ among the results of the three areas, where teachers of the Sociomedical area received evaluations below the Physiological and Morphological areas.

Figure 1 shows the results of the average evaluation of teaching performance by students' opinion, self-assessment and students' academic achievement by area of knowledge. The x-axis shows the three areas and the y-axis the average of the assessments. The white circles represent teachers' self-assessment; black squares teacher evaluation by students' opinion, and the gray diamonds, the academic achievement average of the students. The bars represent the average evaluations of all areas: white=self-assessment (SA), black=students' opinion (SO) and gray= academic achievement (AA).

The ANOVA test showed significant differences among the three areas in relation to the students' academic achievement ($p \leq 0.001$) and self-assessment ($p \leq 0.001$). The comparison of the differences among the general averages of the three teaching performance assessment strategies had a significance of $p \leq 0.001$. This difference becomes more noticeable when we compare self-assessment with students' academic achievement.

Figure 1: Teaching performance assessment by three strategies globally and by area of knowledge.

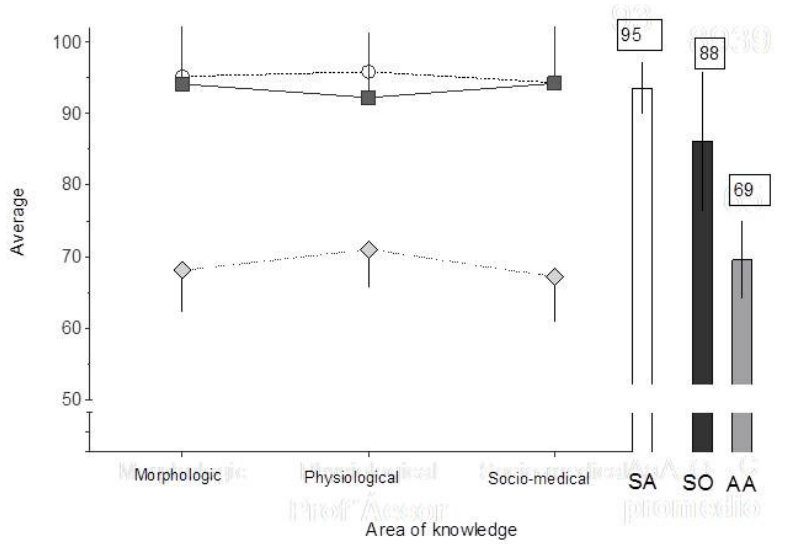



Table 1 shows the statistics by area: average at admission (EES, column A); at graduation (AGES, column B); the differences between them (B-A, Column C) and the adjusted average at graduation (adjusted AGES, column D); the differences between average at entrance and adjusted average at graduation (C-A, Column E), as well as the significant differences by Student's t test. A total of 6,290 scores obtained by students were assessed.

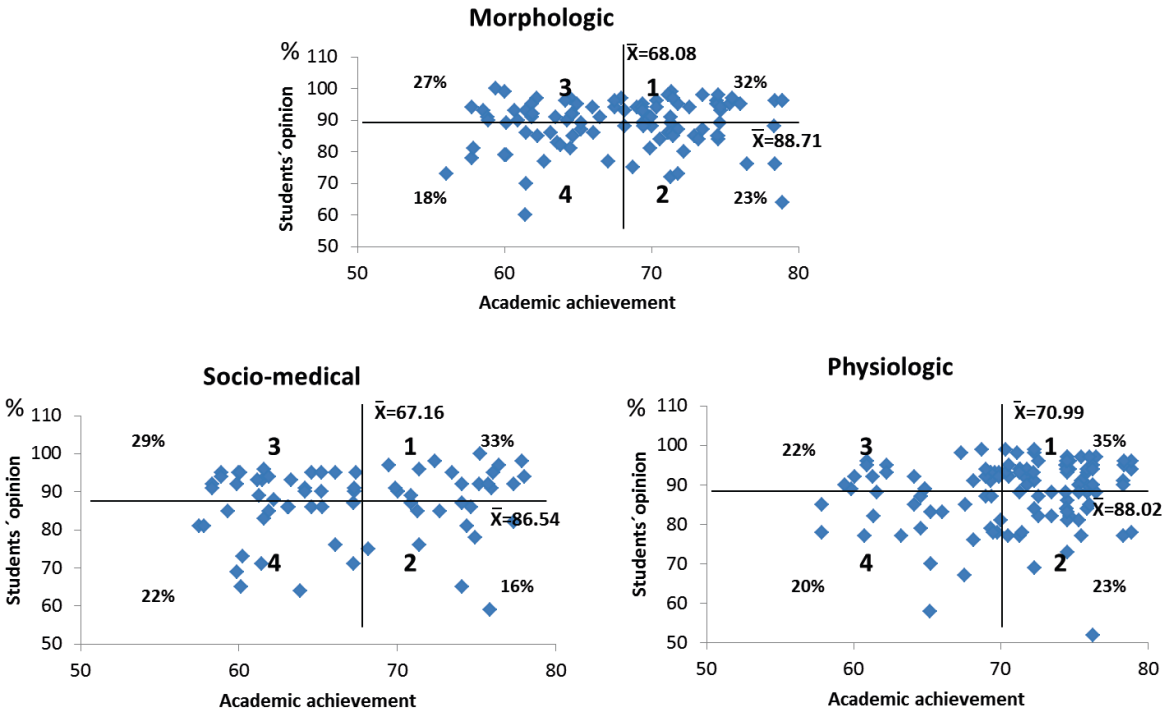
Table 1. Averages at admission and graduation by area

| Area | n | A | B | C | D | E | T-student |
|---|---|---|---|---|---|---|---|
| Morphologic | 2652 | 58.5±13.4 | 68.2± 10.8 | 9.61 | 68.1±5.6 | 9.5 | 0.001 |
| Physiological | 2226 | 65.2±12.1 | 66.1± 12.2 | 0.88 | 71.0±5.3 | 5.7 | 0.001 |
| Sociomedical | 1412 | 56.4±14.4 | 75.1 ±9.8 | 18.67 | 67.2±6.3 | 10.7 | 0.001 |

Figure 2 shows the result of charting by area the averages of group academic achievement and assessment of teaching performance by student's opinion; we divided the sample considering the average of academic achievement and student's opinion. It also shows the classification into four quadrants of teaching performance, corresponding to the following levels:

1. High performance, scores above average in both students' opinion and their academic achievement.
2. Medium High, scores above average in academic achievement, but below average in students' opinion.
3. Medium Low, scores below average in academic achievement, but above average in students' opinion.
4. Low performance, scores below average in both academic achievement and students' opinion.

Figure 2. Classification of teachers by area, according to academic achievement average and students' opinion of teaching performance.




In each area the percentages of teachers who were classified in each quadrant are shown, the lines divide the x and y axes according to the students' opinion and their academic achievement.

In Table 2 we can see the number and percentage of teachers by subject and area according to the level of teaching performance. In the high performance category, the course of Microbiology and Parasitology stands out.

**Table 2 Professor distribution by quadrant according to area and subject taught**

| | Number (%) of professors according to their classification | | | | | |
|---|---|---|---|---|---|---|
| **Subject by area** | High performance | Medium high | Medium low | Low performance | $X^2$ square | P value |
| **Morphologic** | | | | | | |
| Anatomy | 3(3) | 4(4) | 14(14) | 10(10) | | |
| Cell Biology | 3(3) | 1(1) | 7(7) | 5(5) | | |
| Developmental Biology | 3(3) | 0 | 7(7) | 3(3) | | |
| Microbiology and Parasitology | 24(23) | 18(17) | 0 | 0 | | |
| SUBTOTAL | 33(32) | 23(22) | 28(28) | 18(18) | 4.902 | 0.179 |
| **Physiological** | | | | | | |
| Biochemistry | 2(2) | 1(1) | 13(11) | 19(16) | | |
| Pharmacology | 13(11) | 7(6) | 3(3) | 2(2) | | |
| Physiology | 15(13) | 16(13) | 7(6) | 0 | | |
| Immunology | 10(8) | 3(3) | 3(3) | 2(2) | | |
| SUBTOTAL | 40(34) | 27(23) | 26(23) | 23(20) | 5.862 | 0.144 |
| **Socio.medical** | | | | | | |
| Medical Psychology | 5(7) | 4(6) | 10(15) | 10(13) | | |
| Public Health I | 5(7) | 2(3) | 9(13) | 6(9) | | |
| Public Health II | 13(19) | 5(7) | 1(1) | 0 | | |
| SUBTOTAL | 23(33) | 11(16) | 20(29) | 16(22) | 4.629 | 0.201 |

Pearson correlation was significant ($p \leq 0.05$) with a value of 0.248 between students' academic achievement and the teaching strategies used by the teacher (one of the three factors considered in the evaluation by students' opinion).

## 4. Discussion

A comprehensive assessment of teachers is a great challenge since it must be fair and objective, as well as useful, to improve the educational process in medical schools. The lack of a solid evaluation culture in our society prevents the acknowledgement of teachers with a high-quality teaching performance.

The results of this study show that teachers' judgment about their own performance compared to students' perception are generally overestimated, especially in the Physiological area (Figure 1). The scores of students' opinion have a higher standard deviation when compared to those of teacher's self-assessment, which allows the detection of variability in teaching performance that exists among teachers according to students; a difference that may be due to the students' greater objectivity in assessing compared with that of teacher's self-assessment since the latter can be influenced by subjectivity and self-complacency, as reported by other authors (Rippey, 1981; Solabarrieta, 1996; Forest & Chip, 2004). This divergence between the two assessment strategies relies on the premise that students and teachers have different conceptions of what defines an "ideal teacher". Das and Bener (1996) reported that the ability to transmit knowledge is considered by students to be the most important characteristic of a teacher; two additional characteristics mentioned were: teaching material and his/her ability to identify students who do not understand a topic. Nevertheless, teachers have a different perception of what is a good educator; for them the most important qualities are the ability to engage the student in the subject, to promote independent self-learning and to create an appropriate atmosphere for discussion.

In another study in the clinical area, with students and teachers, there were differences ($p \leq 0.001$) in the methods of teaching. As for procedures, they are more important for teachers than for students. Another perception that differed between them was the use of teaching strategies. According to teachers these were frequently used, while students responded that their teachers did not use them so frequently. Hence, teachers are a group prone to give a self-assessment score higher than that given to them by students (Wolf & Turner, 1989), findings similar to this study.

The students' academic achievement strategy is more objective at assessing teaching performance, as it is quantitative (Werther & Keith, 2002), and mathematical models were used to adjust the academic achievement average, considering the average at entrance which involves the degree of knowledge acquired by students; although this does not rule out the great usefulness of students' opinion and, to a lesser degree, teachers' self-assessment.

If we consider the relationship between students' academic achievement and their opinion about teaching, one third of the teachers have high performance in each area. They are above average in assessment by students' opinion and in academic achievement. An important facet of this paper is that it identified, by objective indicators, teachers with high performance, an uncommon situation equivalent to recognizing teaching of outstanding quality.

It is also very important to consider the planning of actions that will immediately strengthen trainee teachers who were evaluated below average in each area (Figure 2, quadrant 4), 22% in the socio-medical area, 20% in Physiological, and also those who need updating in their discipline (located in quadrant three of the graphs by area: 20% in the Physiological area, 22% in Socio-medical, and 27% in the Morphological).

In a study reported by Elizalde and Reyes (2008), 28 dimensions related to teacher effectiveness were identified. They were grouped according to their degree of importance, among which those that had to do with another idea related to the scoring in teaching strategies (clarity and understanding, fostering interest in the subject among students, the perceived impact of training, preparation and course organization by the teacher, fulfilment of objectives for the course and motivation of students) were highlighted.

Human beings need to work with ability, interest and values, because when the relationship between individual skills and the assigned work is balanced, the work becomes a powerful growth factor that promotes development (Garcia, 2010) as shown in Figure 2, right upper quadrant (marked with number 1).

A limitation of this study is that the results represent a sample that could be extrapolated to the entire population of teachers who teach in the first and second medical years at UNAM Faculty of Medicine, but not to populations of teachers from other medical schools or faculties. It should also be considered that in assessing students' academic achievement only the level of knowledge obtained was considered, while skill and ability acquisition by students in relation to the teachers' performance remains an issue for further evaluation.

## 5. Conclusion

It is important for universities to have systematic, comprehensive and objective teaching performance evaluations,

not only for institutional quality improvement, but also for teacher development. Through the combination of two objective assessment strategies, this study identified high performing teachers, those who require training in some teaching skills (teachers rated medium-high and medium-low performance), and those who are not fulfilling their teaching responsibilities.

This line of research must continue in order to learn what and how teachers rated as high performance actually do in their teaching activities, since they represent a role model to follow in the Faculty of Medicine; furthermore, there is a need to properly acknowledge the high-level teaching in our institution.

## 6. Acknowledgments

The authors thank Josefina Bolado for the manuscript´s English review.

## 7. References

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics,* 29, 37-65.

Barbier, M. (1993). *La evaluación en los procesos de formación*. Barcelona: Paidós y Ministerio de Educación y Ciencia.

Cruz, T. (1985). Desempeño académico del profesor del curso de Histología de la Facultad de Medicina, UNAM. *Revista Facultad Medicina UNAM,* 28, 116-123.

Das, M., Sabban, F., & Bener, A. (1996). Student and faculty perceptions of the characteristics of an ideal teacher in a classroom setting. *Medical Teacher*, 18, 141-146.

Elizalde, L., & Reyes, R. (2008). Elementos clave para la evaluación del desempeño de los docentes. *Revista Electrónica de Investigación Educativa*, Especial. Retrieved from http://redie.uabc.mx/NumEsp1/contenido-elizaldereyes.

Fernandez, J. (1997). Spanish evaluation model in higher education: Circular evaluation. *Higher Education Management, 9*, 71-84.

Forest, J., & Chip, H. (1996). The Evaluation Gap in Performance Perceptions: Illusory Perceptions of Groups and Individuals. *Journal of Applied Psychology*, 81 (4), 369-379.

García, T. (2010). Los valores humanos y su relación con el desempeño laboral. (Tesis de Doctorado). UNAM. México.

García-Carrasco, J. (1984). *Teoría de la Educación. Diccionario Ciencias de la Educación*. Madrid, España: Ed. Anaya.

Gimeno-Sacristán, J., & Pérez-Gómez, A. (1989). *La enseñanza: su teoría y su práctica.* Madrid, España: Ed. AKAL.

Ibáñez-Martín, J.A. (1990). Dimensiones de la competencia profesional del profesor universitario. *Rev Esp Pedag*, 186, 239-257.

Irby, D. (1983). Evaluating instruction in medical education. *Journal of Medical Education, 58*, 844-849.

Lancaster, C.J., Ross, G.R., & Smith, I.K. (1988). Survey of Practices in Evaluating Teaching in U.S. Medical Schools, 1978 and 1986. *Journal of Medical Education*, 63, 913-914.

Luna, E., & Torquemada, A. (2008). Los cuestionarios de evaluación de la docencia por los alumnos: balance y perspectivas de su agenda. *Revista Electrónica de Investigación Educativa,* Especial. Retrieved from: http://redie.uabc.mx/NumEsp1/contenido-lunatorquemada.html

McCaffrey, D.F., Lockwood, J.R., Koretz, D.M., & Hamilton, L.S. (2003). Evaluating Value-Added Models for Teacher Accountability. Santa Mónica, CA: The RAND Corporation.

Marsh, H. (1984). Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and utility. *Journal of Educational Psychology*, 76(5), 707-754.

Marsh, H., Fleiner, H., & Thomas, C. (1975). Validity and usefulness of student evaluations of instructional quality. *Journal of Educational Psychology*, 67(6), 883-889.

Martínez-González, A., Gil-Miguel, A., Rey-Calero, J., Cabrera, V. A., Ponce, R. R., & Rodríguez, R. (1999). La evaluación diagnóstica como mecanismo para estimar el éxito o el fracaso en la Licenciatura de Médico Cirujano. *Revista de Educación Superior ANUIES*, 28,5-19.

Martínez- González, A., Fortoul van der Goes, T. & Urrutia- Aguilar, M.E. (2011). Evaluación del desempeño docente en Biología Celular y Tisular mediante tres estrategias. *Archivos en Medicina Familiar*, 13 (2), 101-110.

Martínez González, A., López Barcenas, J., Herrera Saint-Leu, P., Ocampo Martínez J., Petra I., Uribe Martínez G., García Sahagún M.C., & Morales López S. (2008). Modelo de competencias del profesor de medicina. *Educación Médica*, 11(3), 157-67.

Martínez- González, A., Moreno-Altamirano, l., Ponce-Rosas, E.R., & Urrutia-Aguilar, M.E. (2011). Evaluación del desempeño docente en Salud Pública mediante tres estrategias. *Gaceta Médica de México*, 147 (3), 234-243.

Micu, I., Castillo, R., & Díaz, M. (1986). Factores de no acreditación referidos por los alumnos afectados. *Revista Facultad Medicina UNAM,* 29(6), 251-256.

Mullan, B.P., Wolf, F.M., & Ertel, K.I. (1989). Residents' evaluation of behavioral pediatrics instruction. *Journal of Medical Education*, 23, 447-452.

Rodríguez, R., Martínez, A., Ponce, R., Contreras, E., Colina, C., & Cerritos, A. (2000). Una nueva estrategia para evaluar la calidad del desempeño docente en las instituciones de educación superior. *Revista de Educación Superior ANUIES*, 115, 129-141.

Rippey, R.M. (1981). *The evaluation of teaching in medical schools.* New York: Springer Publishing.

Sanmartí, N., & Alimenti, G. (2004). La evaluación refleja el modelo didáctico: análisis de actividades de evaluación planteadas en clases de química. *Educación Química*; 15 (2), 120-128.

Slotnick, B.H., Rusell, G., & Durkovich, G. (1975). Dimensions of medical students' perceptions of instruction. *Journal of Medical Education*, 50, 662-666.

Solabarrieta, J. (1996). Modelos de evaluación del profesor. In Tejedor, F. J. y Rodríguez Diéguez, J. L. (Eds.), *Evaluación Educativa II. Evaluación Institucional. Fundamentos teóricos y aplicaciones prácticas*. Documentos didácticos 157, Salamanca: IUCE. Universidad de Salamanca.

Talayero, U., Petra, M., & Aguilar, G. (1988). Importancia de las actitudes del profesor. *DIDAC*, 13, 38-40.

Téllez-Villagra, C., García, R. H., Nevarez, E., Gallegos, N., and Santana, C. (1990). Factores de riesgo en el desempeño académico deficiente. *Revista Facultad Medicina*, *UNAM*, 33(4), 235-240.

Urrutia, M.E., Barojas, J., Torres, J.R., Ponce, E.R., & Martínez, A. (2008). Predictores de autoevaluación del desempeño docente en ciencias de la salud. *Educación en Ciencias de la Salud*, 5(2), 21-30.

Valle, R., Alaminos, I., Contreras, E., Salas, L., Tomasini, P., & Varela M. (2004). Student Questionnaire to Evaluate Basic Medical Science Teaching (METEQ-B). *Revista Medica IMSS*, 42 (5), 405-411.

Vega-Vega, J.L. (1986). *Psicología de la Educación. Diccionario Ciencias de la Educación.* Madrid, España: Ed. ANAYA.

Werther, W. & Keith, D. (2002). Administración de personal y recursos humanos. México: Interamericana.

Wilson, R. (1986). Improving faculty teaching. *Journal of Higher Education*, 57, 196-211.

Wolf, F., & Turner, E. (1989) Congruence between student and instructor perceptions of clinical teaching in pediatrics. *Medical Education*, 23, 161-167.